Siyue Tian

tian.siyue@northeastern.edu linkedin.com/in/siyue-tian
github.com/JetiGithub

Applied Mathematics M.S. candidate with dual degrees in Mathematics and Statistics and practical Python/R experience in predictive modeling, data preparation, and statistical analysis. Compares regression and classification methods, tunes models, and communicates evaluation results through documented notebooks and visual diagnostics.

Education

Northeastern University, Boston, MA Sept. 2025 – Dec. 2027 (expected)
M.S. in Applied Mathematics; Concentration in Data Science
Selected coursework: Machine Learning II; Numerical Analysis I; Probability II (in progress).

University of California, Davis, Davis, CA Sept. 2021 – Dec. 2024
B.S. in Mathematics and B.S. in Statistics
Selected coursework: Statistical Learning I; Multivariate Data Analysis; Applied Time Series Analysis.

Technical Skills

Programming: Python, R, SQL, MATLAB
Machine learning: scikit-learn, XGBoost; regression, classification, clustering, cross-validation, grid search; RMSE, MAE, R^2, residual analysis
Statistics / modeling: Hypothesis testing, confidence intervals, bootstrap methods, regularization, probability simulation
Data / visualization: pandas, NumPy, dplyr; data cleaning, categorical encoding, feature engineering; Matplotlib, Seaborn, ggplot2
Tools: Jupyter Notebook, R Markdown, SQLite; documented analysis scripts and reports

Selected Machine Learning and Data Science Projects

Predictive Modeling of Boston Airbnb Listings Spring 2026
Northeastern University | Python, scikit-learn, XGBoost

- Transformed 4,419 raw listings into 3,466 analytical records through price conversion, missing-value handling, filtering, and log transformation; constructed 63 encoded predictors.
- Compared Random Forest, Gradient Boosting, XGBoost, and ridge regression on a 694-record test split using RMSE, MAE, and R^2; tuned Gradient Boosting with five-fold GridSearchCV.
- Produced feature-importance and residual plots alongside model-comparison tables; documented preprocessing choices and results in a Jupyter notebook and final report.

Vehicle Insurance Claim Modeling Spring 2024
UC Davis | Team project; Python, R, scikit-learn

- Analyzed a 58,592-record vehicle-insurance dataset, separating numeric, categorical, and binary predictors for exploratory analysis and claim classification.
- Contributed preprocessing and exploratory modeling to a project with Random Forest classification and imbalance-handling experiments; examined feature importance, correlations, and dimensionality reduction in R.
- Documented model experiments and analytical outputs in notebooks and a project report, connecting vehicle and customer attributes to the claim-prediction task.

Least-Squares Stability and Computational Cost Spring 2026

Northeastern University | Python, NumPy, Matplotlib

- Benchmarked five least-squares approaches on simulated $2{,}000 \times 500$ matrices, comparing normal equations, explicit inversion, QR, SVD, and NumPy least squares.
- Evaluated solution error under controlled conditioning, noise, float32/float64 precision, and ridge regularization; visualized numerical behavior and computational cost.

Data Preparation and Statistical Exploration Fall 2025

Northeastern University | R, dplyr, R Markdown

- Cleaned a modified obesity teaching dataset with group-specific height and age imputation; filtered weight outliers, retained 2,094 records, and derived BMI.
- Calculated confidence intervals, category summaries, and overall and stratified correlations; recorded transformations and findings in an analytical notebook.

Image Reconstruction Benchmark Spring 2026

Northeastern University | MATLAB, Python, NumPy, Pillow

- Compared nearest-neighbor, bilinear, bicubic, and Lanczos interpolation on photographs and crops; evaluated multiple downsampling factors with RMS error, PSNR, and edge-region diagnostics.
- Generated comparison figures and CSV outputs and presented how interpolation choice affects reconstructed detail and numerical error.

Professional and Community Experience

Beijing Yintong Chuangda Technology Co., Ltd. Beijing, China

Network Engineering Intern – Data Analytics & AI | Jun. 2023 – Sept. 2023

- Analyzed supply, inventory, and sales data to support replenishment decisions and identify opportunities to reduce excess stock.
- Supported AI-assisted analytics workflows and contributed to an AI-driven merchandise and business-performance analysis platform.

DataLab – maptimeDavis, UC Davis Jan. 2024 – Apr. 2024

Geospatial Data Analysis | Community Projects

- Used QGIS and spatial SQL to map Davis-area geographic data and existing bee-movement data around Davis and the San Francisco Bay Area; communicated spatial patterns through maps.

Think Academy Bay Area, CA

Math Instructor | Jan. 2025 – Sept. 2025

- Taught AMC 8 and AMC 10 preparation to approximately 100 students across five classes; used grading, mock-exam review, and progress tracking to guide instruction.